FLORe – Florence Repository. L'Archivio ad accesso aperto dell'Università di Firenze

Gaia Innocenti, Cristina Mugnai, Valdo Pasqui

1 Introduzione

Anche in Italia l'accesso aperto alle pubblicazioni scientifiche è diventato ormai un tema di pressante attualità sia dal punto di vista normativo che per quanto riguarda gli aspetti organizzativi e le ricadute operative nel mondo della ricerca. L'open access non è più, come in anni ancora recenti, una chimera, la visione dei bibliotecari e dei ricercatori di alcuni ambiti disciplinari, ma si tratta di una realtà in crescita sorretta dalla maggiore consapevolezza dei vantaggi che ne derivano da parte degli attori coinvolti e supportata da linee d'indirizzo che insieme concorrono a realizzare uno scenario impensabile e insperato fino a pochi anni fa. La Commissione Europea ha sancito che l'accesso aperto alle pubblicazioni scientifiche è uno dei principi del nuovo Framework Programme for Research and Innovation, Horizon 2020, la cui portata si estende dal 2014 al 2020 e nel quale è stato stabilito che a partire dal 2014 tutti gli articoli prodotti grazie al finanziamento di fondi del programma dovranno essere accessibili secondo una delle due modalità consolidate:

JLIS.it. Vol. 5, n. 1 (Gennaio/January 2014).
DOI: 10.4403/jlis.it-9102

a. ″gold″ open access—articoli immediatamente resi accessibili online dall'editore;

b. ″green″ open access—i ricercatori rendono i loro articoli disponibili attraverso repository ad accesso aperto entro 6 mesi (12 per gli articoli negli ambiti delle scienze sociali e umanistiche).

In un comunicato stampa del 17 luglio 2013 la Commissione Europea ha ribadito questa impostazione raccomandando agli «Stati membri di adottare un approccio simile anche per i risultati della ricerca finanziata con i programmi nazionali. L'obiettivo è rendere accessibile entro il 2016 il 60% degli articoli scientifici su lavori finanziati con fondi pubblici europei secondo la modalità dell'accesso aperto»e inoltre la Commissione ha espresso la volontà di iniziare anche a «sperimentare l'accesso aperto con i dati raccolti nel corso delle ricerche finanziate con fondi pubblici».[1] Nel mese di agosto 2013 è stato pubblicato uno studio (Archambault et al.) condotto da Science-Metrix per conto della Direzione Generale Ricerca & Innovazione della Commissione Europea che ha esaminato 22 ambiti di conoscenza su un campione di paesi comprendente, oltre ai 28 dell'Area Europea, anche Svizzera, Norvegia, Islanda, Brasile, Canada, Giappone, Stati Uniti ed altri. Lo studio afferma che in molti di questi paesi e per parecchie aree disciplinari ormai più del 50% degli articoli sono disponibili gratuitamente e che addirittura in discipline come la tecnologia, la ricerca biomedica, la biologia, la matematica e la statistica la maggioranza dei lavori sono disponibili ad accesso libero. In Italia, dopo le azioni di sensibilizzazione condotte dalla Commissione Open Access della CRUI[2] diversi Atenei hanno inserito nei propri statuti impegni precisi rispetto alla promozione dell'accesso aperto. Fatto del tutto nuovo ed in parte imprevisto, il 10 agosto è entrato in vigore il Decreto Legge 8

[1] http://europa.eu/rapid/press-release_IP-12-790_it.htm.
[2] http://www.crui.it/HomePage.aspx?ref=894.

agosto 2013, n. 91, recante "Disposizioni urgenti per la tutela, la valorizzazione e il rilancio dei beni e delle attività culturali e del turismo" il cui Art.4 comma 2 prevedeva che «Le pubblicazioni che documentano i risultati di ricerche finanziate per una quota pari o superiore al cinquanta per cento con fondi pubblici, indipendentemente dal formato della prima pubblicazione e dalle modalità della sua distribuzione o messa a disposizione del pubblico, devono essere depositate, non oltre sei mesi dalla pubblicazione, in archivi elettronici istituzionali o di settore, predisposti in modo tale da garantire l'accesso aperto, libero e gratuito, dal luogo e nel momento scelti individualmente, l'interoperabilità all'interno e all'esterno dell'Unione Europea e la conservazione a lungo termine in formato elettronico». Successivamente, in sede di conversione in legge, la Commissione cultura del Senato ha ampiamente emendato il testo attenuandone molto il dispositivo iniziale. Infatti la formulazione finale, approvata dai due rami del Parlamento, recita: «I soggetti pubblici preposti all'erogazione o alla gestione dei finanziamenti della ricerca scientifica adottano, nella loro autonomia, le misure necessarie per la promozione dell'accesso aperto ai risultati della ricerca finanziata per una quota pari o superiore al cinquanta per cento con fondi pubblici, quando documentate in articoli pubblicati su periodici a carattere scientifico che abbiano almeno due uscite annue. L'accesso aperto si realizza: a) tramite la pubblicazione da parte dell'editore, al momento della prima pubblicazione, in modo tale che l'articolo sia accessibile a titolo gratuito dal luogo e nel momento scelti individualmente; b) tramite la ripubblicazione senza fini di lucro in archivi elettronici istituzionali o disciplinari, secondo le stesse modalità , entro 18 mesi dalla prima pubblicazione per le pubblicazioni delle aree disciplinari scientifico-tecnico-mediche e 24 mesi per le aree disciplinari umanistiche e delle scienze sociali». Pur nella profonda revisione del contenuto la disposizione convertita in

legge deve essere accolta come il segnale che l'Italia ha comunque iniziato a recepire dal punto di vista normativo gli orientamenti europei e che è pronta a percorrere in modo più formalizzato e concreto la strada dell'accesso aperto. In questo contesto l'Università di Firenze è stata un'antesignana poiché già nell'anno 2000, tra i primi atenei in Italia, ha avviato la sperimentazione di un repository open access promuovendo il "self archiving" ("green road") con un progetto all'epoca pionieristico perché nato senza un quadro di riferimento tecnologico, bibliografico (i metadati), organizzativo e normativo consolidato e pertanto volto anche a valutare e sperimentare sia le possibili soluzioni tecniche che i modelli organizzativi e di comunicazione più adatti per promuovere la "cultura" dell'accesso aperto.

2 L'open access all'Università di Firenze

2.1 L'esperienza pregressa

Il primo archivio open access dell'ateneo fiorentino, denominato Archivio E-Prints, è stato realizzato oltre 10 anni fa nell'ambito del progetto per la costituzione della biblioteca digitale promosso dalla Firenze University Press (FUP). In via preliminare uno studio svolto nella primavera del 2000 (Pasqui) permise di delineare il framework di riferimento articolato, sulla base delle visioni e delle prospettive di quegli anni, secondo tre principi:

a. preservare l'autonomia, sia dal punto di vista gestionale che da quello dei contenuti, delle iniziative esistenti e future di facoltà, dipartimenti e centri;

b. garantire l'interoperabilità tra gli archivi in corso di costituzione;

c. consentire agli utenti finali l'accesso integrato alle risorse, in particolare al catalogo in linea delle biblioteche (OPAC), alle collezioni digitali primarie e ai *subject information gateways*.

Tenendo conto di questi principi e del contesto funzionale (tipi utente, tipologie dei materiali, modalità di auto-archiviazione) lo studio condusse all'individuazione di un modello basato sulla Open Archives Iniziative (OAI) assumendo in particolare l'articolazione in Data Provider e Service Provider e il protocollo Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Per la predisposizione dell'archivio, tra i sw open source all'epoca disponibili, fu scelto il software eprints.org sviluppato dall'Università di Southampton. L'archivio fu avviato nel gennaio 2001. Dopo una fase di sperimentazione, durata circa un biennio, il repository raggiunse uno stadio di consolidamento attraverso alcuni interventi di personalizzazione e configurazione del sw eprints.org quali lo sviluppo di una versione bilingue (italiano/inglese) e l'implementazione dello schema di soggettazione conforme ai settori scientifico disciplinari (14 Aree) definiti dal Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR)[3]. La registrazione presso l'harvester mondiale OAIster garantiva la possibilità di promuovere la disseminazione della produzione scientifica Open Access (OA) dell'ateneo. Dal punto di vista organizzativo e operativo le politiche di governo dei contenuti e dei rapporti con gli autori erano gestite dalla FUP, mentre il Centro Servizi Informatici dell'Ateneo Fiorentino (ora Sistema Informatico dell'Ateneo Fiorentino (SIAF)) garantiva lo sviluppo e la gestione tecnologica. Nel febbraio del 2004, in occasione del convegno organizzato per la presentazione ufficiale dell'archivio, alcuni interventi misero in luce aspetti che a distanza di anni si sarebbero rivelati premonitori delle criticità manifestatesi per questa iniziativa. Giancarlo Pepeu, allora Presidente della FUP, nella introduzione

[3] http://www.miur.it/UserFiles/115.htm.

affermava che «qualsiasi archivio e-prints deve dare garanzie sulla correttezza e l'attendibilità di quanto è in esso contenuto»e «se non vi è una garanzia della validità di quanto si trova negli archivi e-prints, essi diventeranno depositi di informazioni dubbie, di scarso valore, ai quali nessuno attingerà»(Pepeu). Patrizia Cotoneschi, nel porre in risalto le sinergie dell'archivio con l'attività editoriale, sottolineava che l'archiviazione non è «un'alternativa alla pubblicazione tradizionale, ma un mezzo complementare per diffondere i risultati della ricerca»e annunciava la predisposizione di un nuovo tipo di contratto FUP che avrebbe contemplato per gli autori «la possibilità di diffondere liberamente il formato elettronico delle loro pubblicazioni»(Cotoneschi). Infine, Giulia Maraviglia, dopo aver ricordato la consistenza dell'archivio (116 contributi per un totale di 56 autori), rilevava che la

> oggettiva lentezza con cui sta crescendo l'archivio e-print dell'ateneo fiorentino può essere quasi certamente attribuita alla scarsa fiducia che la comunità accademica dimostra ancora di avere verso questa iniziativa. Ai bibliotecari tuttavia spetta il compito di promuovere e diffondere l'utilizzo di questo strumento, spesso per molti ancora sconosciuto. L'auspicio è che il deposito e la messa a disposizione dei materiali sopra elencati diventi prassi di lavoro in tutti gli atenei (Maraviglia).

Nonostante l'Università di Firenze il 6 giugno 2005 avesse aderito alla Dichiarazione di Berlino per l'accesso aperto alla letteratura scientifica[4], negli anni successivi è venuta meno la capacità di intraprendere azioni in grado di promuovere efficacemente il deposito della produzione scientifica nell'archivio e di assicurarne uno sviluppo sistematico e coerente. Per queste criticità è stata determinante

[4] http://openaccess.mpg.de/286432/Berlin-Declaration.

soprattutto l'assenza di una policy che definisse gli attori, i principi e gli obblighi, fatto che ha fortemente limitato lo sviluppo del self-archiving. Di conseguenza l'archivio è rimasto uno strumento di nicchia per pochi autori, con un contenuto molto eterogeneo concernente non solo la produzione scientifica, incluse alcune tesi di dottorato, ma anche supporti didattici di vario tipo (dispense e slide) ed altro materiale per una consistenza che alla fine del 2011 superava di poco i 1200 documenti, con un numero di nuovi contributi per anno inferiore a 100 eccetto pochi anni (2004, 2005 e 2007). Ovviamente il limitato utilizzo dell'archivio e-prints ne ha determinato anche l'obsolescenza tecnologica, pur non inficiandone il funzionamento. Infatti è stato inevitabile dedicare le risorse umane disponibili in possesso di competenze informatiche ai progetti dell'ateneo con più elevata priorità, spesso finanziati con contributi esterni (fondazioni bancarie, MIUR, Presidenza del Consiglio dei Ministri), relativi ai servizi online per studenti (es. la verbalizzazione degli esami con la firma digitale, la valutazione della didattica, il fascicolo elettronico dello studente etc). Nell'ultimo biennio, come vedremo tra breve, il riesame delle prospettive e la rivalutazione delle strategie relative all'accesso aperto all'interno dell'Università di Firenze hanno portato a sviluppare e realizzare una nuova soluzione per il repository istituzionale dell'ateneo.

2.2 La genesi di Florence Repository (FLORe)

Sulla base dall'esperienza pionieristica, ma dagli esiti non soddisfacenti, dell'archivio e-prints descritta nel paragrafo precedente e grazie all'opera di advocacy a favore dell'accesso aperto condotta da Mauro Guerrini nel maggio del 2011 il rettore dell'Università Firenze ha istituito la commissione di ateneo "sul tema accesso aperto per i prodotti della ricerca e per le tesi di dottorato", presieduta dal prorettore alla ricerca Elisabetta Cerbai e comprendente alcuni

docenti e ricercatori in rappresentanza delle diverse aree disciplinari, il Sistema Bibliotecario d'Ateneo (SBA), la FUP e il SIAF. La sistematicità e il rigore con i quali la commissione ha condotto la propria attività sotto la guida di Elisabetta Cerbai, decisamente coinvolta nella tematica dell'accesso aperto, hanno portato nel giro di pochi mesi a finalizzare alcuni obiettivi molto concreti. Il primo risultato in ordine temporale e fondamento di tutte le fasi successive è stata la definizione della "Policy di ateneo sull'open access"[5] che, dopo un ampio e approfondito confronto con la Commissione Ricerca, è stata approvata all'unanimità dal Senato Accademico nella seduta del 11 aprile 2012. Successivamente, grazie alla sensibilità del rettore Alberto Tesi, è avvenuto il recepimento dell'impegno a favore dell'accesso aperto nel nuovo Statuto dell'Ateneo, entrato in vigore l'8 maggio 2012, che al comma 2 dell'Art.8 "Ricerca scientifica" recita: «L'Università informa la disciplina delle attività di ricerca ai principi della trasparenza e della pubblicità; fa propri i principi dell'accesso pieno e aperto alla letteratura scientifica e promuove la libera diffusione in rete, nei circuiti della comunità scientifica internazionale, dei risultati delle ricerche prodotte in Ateneo». Infine, il lavoro della commissione si è focalizzato sulla individuazione delle soluzioni più idonee per dotare l'ateneo di un repository istituzionale destinato a divenire lo strumento per promuovere l'accesso aperto ai prodotti della ricerca e in grado di supportare adeguatamente i docenti ed i ricercatori. L'approccio seguito ha privilegiato l'aspetto funzionale rispetto a quello tecnico, ovvero invece di procedere direttamente alla scelta dello strumento informatico attraverso il quale implementare il repository si è data la priorità alla valutazione della situazione esistente e dei requisiti funzionali, evitando così di far prevalere le scelte relative alla piattaforma tecnologica sulle esigenze di carattere funzionale e operativo. L'analisi svolta ha posto in evidenza che

[5] http://www.unifi.it/notiziario/upload/sub/2012_2/policy_open_access.pdf.

il costituendo repository istituzionale avrebbe dovuto soddisfare i seguenti requisiti funzionali:

a. *Integrare il flusso di lavoro dei docenti e ricercatori*. Le attività di alimentazione e gestione del repository nella modalità self-archiving devono essere ben integrate ed armonizzate con i workflow di lavoro abituali degli autori (docenti e ricercatori) senza aggiungere oneri operativi;

b. *Evitare il doppio inserimento dei dati*. Nell'ambito del requisito precedente occorre prioritariamente escludere il rischio di inserimenti plurimi dei metadati e dei full text dei documenti;

c. *Valorizzare U-GOV Catalogo Ricerca*. Per l'Ateneo U-GOV Catalogo Ricerca è uno strumento ormai consolidato e con un contenuto rilevante di prodotti/pubblicazioni registrate che rappresentano il risultato del lavoro della comunità scientifica dell'Ateneo e che comprendono numerosi full-text dichiarati ad accesso pubblico;

d. *Facilitare la ricerca*. L'interfaccia di ricerca e consultazione deve essere semplice ed intuitiva e possibilmente allineata con alcuni servizi ben conosciuti (es. Scopus);

e. *Assicurare la disseminazione dei metadati*. Occorre garantire l'integrazione con i metamotori di ricerca di tipo harvesting basati sul protocollo OAI-PMH;

f. *Rendere omogeneo e coerente il contenuto*. Il repository deve contenere ed esporre le pubblicazioni relative all'attività di ricerca svolta dai docenti e ricercatori dell'Ateneo nei vari ambiti disciplinari.

In particolare l'approfondimento dei primi tre requisiti ha messo in evidenza il ruolo di U-GOV Catalogo Ricerca che è stato attivato

presso l'ateneo di Firenze da aprile 2008 e che nel periodo autunno 2011 – primavera 2012 è stato lo strumento usato per il conferimento, da parte di docenti e ricercatori, dei dati relativi alle proprie pubblicazioni (metadati descrittivi e full text) ai fini della Valutazione della Qualità della Ricerca (VQR) secondo le linee guida dettate dall'Agenzia di Valutazione del Sistema Universitario e della Ricerca (ANVUR). Di conseguenza la modalità di inserimento in U-GOV Ricerca è divenuta in quel periodo familiare alla maggior parte degli autori e il catalogo non si è arricchito soltanto dei metadati, la cui registrazione è stata agevolata dalle possibilità di cattura tramite Digital Object Identifier (DOI) e di importazione da formati standard (e.g. BibText, Endnote, Institute for Scientific Information (ISI), Metadata Object Description Schema (MODS)) forniti dagli editori, ma anche del full text di molte pubblicazioni tra cui diverse rese disponibili ad acceso aperto immediatamente o in modalità differita, con la specifica di una data di fine embargo. Questa realtà ha portato a riflettere sui vantaggi e gli svantaggi di affiancare a U-GOV Ricerca un secondo strumento software per implementare il repository open access. Tra i vantaggi si è notata certamente una maggior flessibilità dal punto di vista della configurabilità dei metadati (per quanto anche U-GOV offra ampie possibilità di specificare attributi descrittivi per ogni tipologia di pubblicazione) secondo un approccio più orientato all'ambito bibliografico e tenuto conto che i software utilizzati per la gestione di repository istituzionali (es. eprints.org, DSpace, Fedora Commons) consentono l'adozione di set di metadati estesi rispetto al Dublin Core non qualificato. Un ulteriore vantaggio offerto da tali software sarebbe stata la possibilità di definire dei work-flow attraverso i quali sottoporre le pubblicazioni depositate tramite self-archiving a uno o più passi di verifica, revisione e integrazione dei metadati da parte di esperti prima di essere rese visibili pubblicamente. Questa modalità avrebbe consentito la

manutenzione qualitativa (correzioni, integrazioni, arricchimento) dei metadati bibliografici da parte di bibliotecari competenti nelle varie aree disciplinari. Tra gli svantaggi di un doppio contenitore è emersa la necessità di costringere gli autori a operare su due interfacce diverse (U-GOV Catalogo Ricerca e il repository) non solo per quanto riguarda il layout grafico, ma soprattutto per quanto concerne le modalità di interazione (sequenza di immissione, controlli e vincoli). D'altra parte, U-GOV consente di svolgere agevolmente le operazioni di self-archiving e, per gli atenei che come Firenze lo hanno adottato, è anche lo strumento primario per mantenere aggiornato il sito docente MIUR e pertanto il suo utilizzo costante è divenuto irrinunciabile ed è entrato a far parte del normale flusso di lavoro dei docenti e dei ricercatori. Un secondo svantaggio emerso nell'analisi dell'adozione del doppio repository è rappresentato dalla duplicazione dei metadati e dei full-text che prima di tutto implicano la maggiore occupazione di risorse hardware. Questa motivazione potrebbe essere ritenuta di scarsa rilevanza visto il costo limitato delle risorse hardware, tuttavia si devono considerare i maggiori costi operazionali Operating Expenditure (OpEx) dovuti al mantenimento e alla gestione di due diverse piattaforme tecnologiche. Inoltre occorre valutare anche l'aspetto funzionale, ovvero il fatto che un doppio contenitore/ambiente avrebbe comportato l'inevitabile disallineamento dei metadati per effetto degli interventi di manutenzione dei metadati da parte degli autori e dei team di manutenzione bibliografica sul secondo repository rispetto a quelli originari di U-GOV Ricerca che sono sincronizzati con la banca dati MIUR. Per quanto riguarda l'arricchimento e la manutenzione dei metadati, tenuto conto delle dimensioni dell'ateneo di Firenze (24 Dipartimenti e oltre 1900 tra docenti e ricercatori strutturati che operano su molti settori disciplinari) e delle scarse risorse umane disponibili, è stata valutata anche la difficoltà di attivare team di

bibliotecari in grado di assicurare con continuità la manutenzione qualitativa dei metadati e curarne costantemente l'arricchimento, previa ovviamente la preventiva interazione con gli autori delle pubblicazioni. In U-GOV l'autore che registra un prodotto è anche il "proprietario" e gestore dei metadati e comunque sono previste figure di utenti privilegiati in grado di intervenire su di essi. Infine, rispetto al requisito dell'omogeneità e della coerenza dei contenuti, la commissione si è orientata verso una specializzazione per archivi dedicati. Così il materiale inerente la didattica (dispense, slide, registrazioni audio e video) è stato fatto confluire sulla piattaforma E-learning, da tempo resa operativa da SIAF con il software open source Moodle. Altre tipologie di dati e documenti non strettamente connessi alla produzione scientifica sono ospitati nelle "Collezioni digitali" curate dallo SBA. FLORe si caratterizza dunque come "il Repository Istituzionale della Ricerca" con la doppia interpretazione dell'acronimo FLORe come "FLOResearch" e "FLORepository".

2.3 FLORe in sintesi

Sulla base dei requisiti e delle considerazioni esposti nel paragrafo precedente per assolvere al proprio ruolo funzionale di repository istituzionale della produzione scientifica dell'Università di Firenze, FLORe è stato concepito come una estensione di U-GOV Catalogo Ricerca. In sostanza FLORe è a tutti gli effetti un Repository Open Access che, dal punto di vista logico, sviluppa pienamente il modello originario, già individuato nel 2000, fondato sulle due componenti, Data Provider e Service Provider, dell'architettura distribuita OAI. Il Data Provider, ovvero il nucleo centrale di FLORe, è U-GOV Catalogo Ricerca che fornisce agli autori l'interfaccia e le funzionalità per l'inserimento e la manutenzione dei metadati, per il caricamento dei full-text e per la loro caratterizzazione come documenti ad accesso aperto, con o senza embargo. In questo modo gli autori conservano

immutato il work-flow a cui sono abituati e viene loro evitata la necessità di utilizzare un secondo strumento per l'immissione dei dati bibliografici e delle pubblicazioni, inoltre si evita ogni duplicazione dei metadati e dei contenuti che sono mantenuti in un unico repository fisico. Infine il Data Provider conserva anche le funzionalità aggiuntive, rilevanti per i docenti e i ricercatori, che sono intrinseche di U-GOV Ricerca ovvero:

i. evitare la digitazione manuale delle descrizioni bibliografiche importandole da formati standard o catturando i record bibliografici dai siti degli editori tramite DOI;

ii. il riversamento automatico delle pubblicazioni immesse o modificate nella banca dati personale del docente/ricercatore grazie all'integrazione con il sito docente MIUR;

iii. il "riconoscimento" dell'autore negli archivi delle risorse umane, operando l'aggancio con i dati di carriera (matricola, ruolo etc.) e garantendo così l'integrazione del catalogo ricerca con le informazioni del dominio del personale dell'Ateneo.

Il Data Provider è dotato dell'interfaccia OAI-PMH che Consorzio interuniversitario per il Calcolo Automatico (CINECA) ha implementato nel corso del 2011-2012 su richiesta di Firenze e di altri atenei come l'Università di Verona (Gabaldo). Tramite tale protocollo avviene l'harvesting dei metadati da parte di Pleiadi, il motore di ricerca nazionale per l'accesso centralizzato alla letteratura scientifica depositata negli archivi aperti italiani, istituiti e gestiti da università ed enti di ricerca, sviluppato dal Consorzio Interuniversitario Lombardo per L'Elaborazione Automatica (CILEA) e gestito da CINECA dopo l'unificazione dei consorzi universitari.[6] Infine Il

[6]http://www.openarchives.it/pleiadi.

Service Provider è costituito dall'interfaccia di ricerca e visualizzazione dei risultati realizzata e mantenuta autonomamente da SIAF. Con il contributo dello SBA e il coordinamento di Mauro Guerrini, in base alle specifiche individuate e ai test effettuati nell'ambito del Sottogruppo Tecnico della Commissione Open Access, le funzioni di ricerca sono state sviluppate per versioni successive raffinando le funzionalità e perfezionando gli algoritmi a partire dalla versione Beta che è stata presentata ufficialmente il 25 ottobre 2012 in occasione della settimana dell'Accesso Aperto 2012. Attualmente la maschera di ricerca (Fig.1) propone quattro macro canali: "Titolo, Abstract, Keywords, Autore", "Titolo, Abstract, Keywords", "Autore", "Titolo". I primi due sono stati concepiti per fornire ricerche immediate "alla Google", ovvero sulle singole parole immesse: in tal modo è possibile indicare nel campo di ricerca parole che appartengono a domini diversi. Nei canali "Autore" e "Titolo" la ricerca avviene invece sulla frase esatta (per esempio cercando Mario Rossi in Autore si trovano solo le pubblicazioni in cui compare la stringa "Mario Rossi" nel campo "Autore"). È possibile utilizzare gli operatori logici AND, OR, NOT se indicati in una frase posta tra virgolette. Per raffinare la ricerca sono disponibili alcuni filtri:

- anno di pubblicazione: consente di selezionare un singolo anno oppure un intervallo di tempo da... a...);
- data inserimento: permette di visualizzare le pubblicazioni inserite in un determinato periodo (es. ultimi 15/30/90/180 gg, anno corrente, anno precedente);
- tipologia: il menù a tendina permette di scegliere il tipo di pubblicazione cui si è interessati (es. contributo su rivista, contributo in volume, tesi di dottorato etc.).

Il motore di ricerca prevede anche l'utilizzo di wild card con effetto di generalizzare la ricerca:

- ″%″ sostituisce uno o più caratteri;
- ″_″ indica una posizione singola, all'interno di una parola, in cui può apparire un qualunque carattere;
- ″,″ consente di trovare i risultati, in base al canale prescelto, in cui è presente almeno una occorrenza di uno dei termini indicati. I risultati sono ordinati in base al numero di termini trovati;
- ″$″ permette di cercare parole che hanno la stessa radice linguistica.

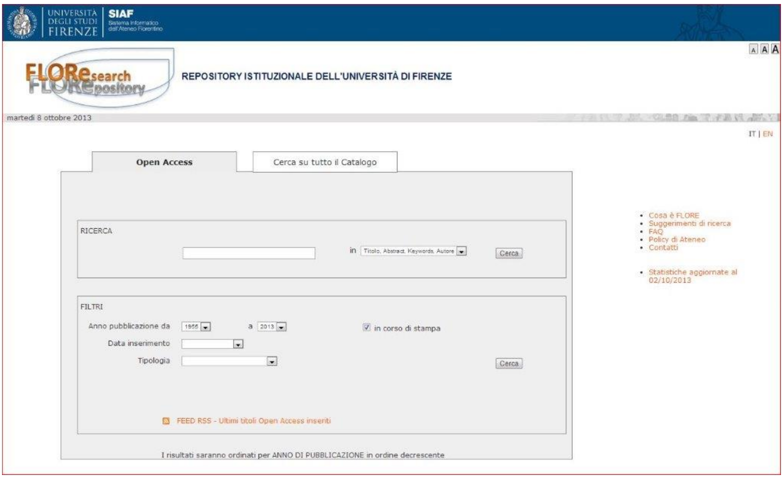


Figura 1: Maschera di ricerca

Se nella stringa di ricerca si specificano parole che non sono indicizzate (stop words, a seconda del canale prescelto e degli eventuali operatori booleani immessi, la ricerca può dare luogo a risultati

nulli. Al momento della stesura di questo articolo è in corso l'assestamento della lista di queste parole e la valutazione se eliminarle completamente. I risultati della ricerca (Fig.2) vengono proposti ordinati per anno di pubblicazione decrescente, quindi mostrando prima le pubblicazioni più recenti, preceduti da icone grafiche che permettono di individuare a colpo d'occhio la tipologia della pubblicazione e sono presentati in pagine successive se il numero supera una soglia prefissata. La descrizione bibliografica sintetica (Fig.3) è cliccabile per consentire la consultazione della scheda bibliografica di dettaglio che propone tutti i campi descrittivi inseriti in U-GOV. Per il full text open access viene mostrato il formato (di solito PDF), la dimensione ed il link "Download" che permette di visualizzare/catturare il documento. Se disponibili i link al sito dell'editore e il DOI vengono sempre visualizzati in conformità a quanto previsto dalle politiche degli editori relative ai diritti di autore[7]. Infine, segue l'indicazione di quali autori sono "Autori dell'Università di Firenze" in base al loro precedente auto-riconoscimento in U-GOV. Da ciascuno di questi è possibile procedere con una ricerca implicita che recupera e propone tutte le pubblicazioni ad accesso aperto dell' autore appartenente all'Ateneo.

Con la stessa modalità ed interfaccia, ma con alcuni filtri aggiuntivi, cliccando sul tab "Cerca su tutto il Catalogo" è possibile effettuare ricerche su tutte le pubblicazioni, ovvero non più esclusivamente su quelle ad accesso aperto. Infine, un brevissimo cenno alla piattaforma tecnologica. I dati bibliografici utilizzati per le ricerche sono derivati da alcune tabelle virtuali (viste) esposte dal database U-GOV opportunamente materializzate e indicizzate con il motore Oracle Text. La soluzione di avvalersi di Oracle non deve destare sorpresa poiché tutti i database del sistema informativo dell'Ateneo, incluso U-GOV, usano il Data Base Management System (DBMS)

[7] http://www.sherpa.ac.uk/romeo.

Oracle; quindi, per ragioni di uniformità ed efficienza oltre che per il know-how esistente, questa è stata la soluzione naturale. Le procedure di interrogazione sono scritte in Java (JSP e servlet) e gestite tramite la piattaforma open source standard Apache-Tomcat, ormai consolidata da oltre 10 anni per lo sviluppo di tutti i servizi online per studenti e docenti. L'accesso ai full-text avviene per mezzo di web-service appositamente esposti da U-GOV: in tal modo i contenuti (file PDF) vengono prelevati e trasferiti solo quando si effettua l'operazione di download.

3 Valutazioni, prospettive e conclusioni

Al 4 novembre 2013 FLORe comprende oltre 6.000 pubblicazioni con full-text open access, di cui 560 sono tesi di dottorato, rispetto al totale di poco più di 129.100 pubblicazioni/prodotti registrate/i in U-GOV Catalogo Ricerca. Dal punto di vista dei metadati, la principale criticità riscontrata è la notevole eterogeneità con la quale sono state registrate le stringhe autore nelle descrizioni bibliografiche. Solo in minima parte è stata usata la forma completa "Nome Cognome"; spesso sono state usate altre forme: iniziale del nome puntata, seguita dal cognome, con o senza uno spazio separatore, oppure il cognome seguito da uno spazio e dall'iniziale del nome puntato. Il separatore delle stringhe autore di solito è la virgola, ma anche per questo vi sono delle eccezioni. Questa varietà di forme può dare luogo a risultati incompleti nella tipica ricerca per autore formulata come "Nome Cognome" (es. Mario Rossi) ed è la conseguenza della difficoltà di assicurare nel tempo, fin dall'avvio del catalogo, uniformità per l'immissione della stringhe autore, tenuto conto del contesto self-archiving, dell'eterogeneità delle fonti bibliografiche e delle consuetudini vigenti nelle diverse aree disciplinari. Valutata l'impossibilità di intervenire ex-post per bonificare ed uni-

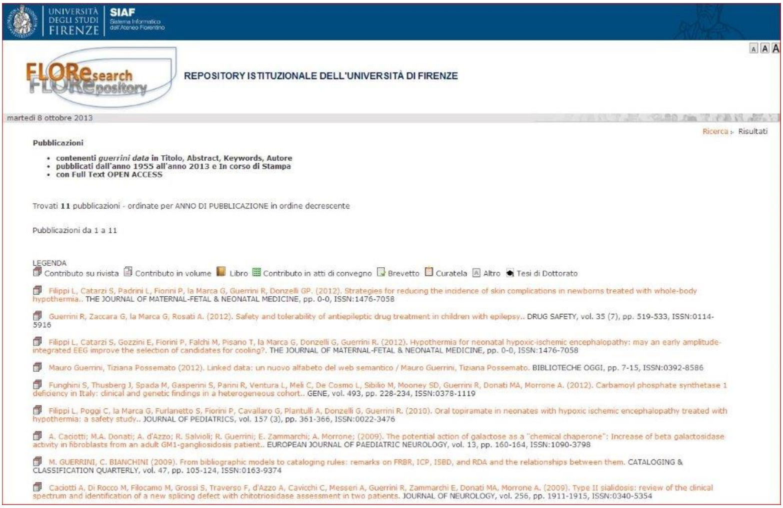


Figura 2: Lista sintetica di risultati

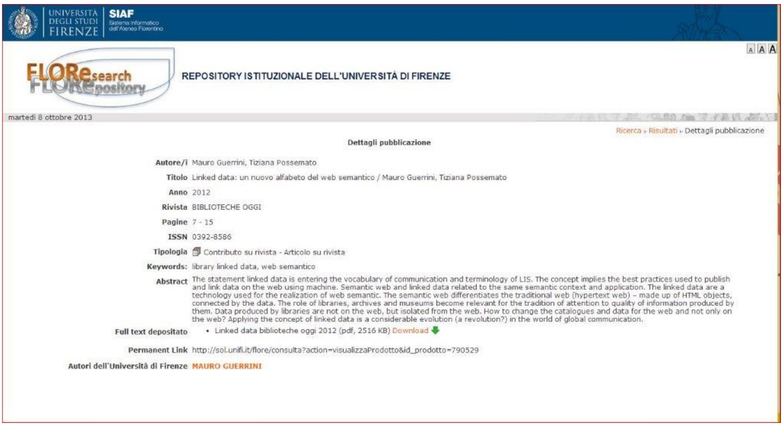


Figura 3: Scheda bibliografica di una pubblicazione

formare le stringhe autore è stata considerata l'ipotesi di restringere la ricerca per autore ai soli autori "riconosciuti" (quelli per i quali, grazie all'aggancio con i dati delle risorse umane che fungono da authority file, sono disponibili il nome ed il cognome corretti), soluzione che è stata poi scartata in quanto non consentirebbe di cercare anche autori esterni all'Ateneo che in molte pubblicazioni di alcune discipline (es. quelle biomediche) sono assai numerosi. A questo limite pone in parte rimedio la funzione di ricerca per "autore riconosciuto" attivabile con un semplice click (link) sulla stringa "Nome Cognome" proposta nella lista "Autori dell'Università di Firenze" a piè di ciascuna scheda bibliografica di dettaglio. È incorso di valutazione anche la possibilità di arricchire la ricerca di FLORe con una funzionalità di browse per "autori riconosciuti" che consenta di recuperare immediatamente per ciascuno di essi le relative pubblicazioni ad accesso aperto. Rientra nelle prospettive a medio termine il processo di convergenza annunciato da CINECA tra U-GOV Catalogo Ricerca e Surplus, il software sviluppato e gestito da CILEA prima della confluenza dei consorzi universitari Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca (CASPUR) e CILEA in CINECA. Al momento della stesura di questo articolo le uniche informazioni disponibili sono quelle relative all'incontro CINECA Library User Group (CLUG) del 9 ottobre[8]. Per quanto riguarda FLORe non vi saranno impatti se, come dichiarato, CINECA manterrà invariata l'esposizione delle viste dei metadati e i web-services di U-GOV Ricerca. Eventuali modifiche architetturali conseguenti all'integrazione di DSpace in U-GOV (es. l'archiviazione dei full-text, tabelle dei metadati etc.) avranno certamente conseguenze a livello tecnico per gli atenei che come Firenze ospitano U-GOV sulle proprie piattaforme, ma non saranno perce-

[8] https://clug.cineca.it/system/files/accesso/riunioni/Webex%20riunione%20Gruppo%20gestione%20dati%20finale_0.doc.

pite dagli utenti finali di FLORe. Diversamente, dal punto di vista funzionale ed operativo, nel caso in cui la convergenza con Surplus determinasse un cambiamento significativo nell'attuale interfaccia grafica di U-GOV Catalogo Ricerca, che è parte integrante di FLORe e della sequenza operativa ormai consolidata per l'inserimento di una pubblicazione e del relativo full-text da parte degli autori dell'ateneo fiorentino, l'impatto sui medesimi autori, docenti e ricercatori, che archiviano le loro pubblicazioni in FLORe sarebbe significativo. In questo caso, quindi, le evoluzioni e le conseguenze dovranno essere attentamente valutate con CINECA. Per quanto riguarda i contenuti del repository, la loro crescita è legata anche alle iniziative che saranno intraprese per promuovere il deposito ad accesso aperto da parte dei docenti e dei ricercatori dell'Ateneo, al fine di assecondare un sempre più ampio spostamento verso l'archiviazione delle pubblicazioni in modalità open access, anche tramite servizi di aiuto e consulenza per la cura dei metadati bibliografici e la verifica delle politiche degli editori. Infine, pur senza voler anticipare soluzioni da verificare e approfondire con più attenzione in futuro, tenuto conto del recente orientamento internazionale e governativo verso la pubblicazione di dati aperti (open data), FLORe è già predisposto anche per consentire la pubblicazione e la diffusione di set di dati relativi a una ricerca scientifica; i relativi file possono infatti essere allegati alla pubblicazione ed essere messi a disposizione in modalità open access secondo formati aperti. Nel contesto dei requisiti individuati e delle prospettive tratteggiate, l'implementazione del repository istituzionale FLORe è una soluzione efficace sia sul piano tecnico che su quello funzionale poiché estende e integra U-GOV Catalogo Ricerca minimizzando l'entità dei moduli applicativi appositamente sviluppati, dunque riducendo il costo gestionale, inoltre non altera le modalità di lavoro da parte dei docenti e dei ricercatori, evitando di distoglierli con altre interfacce e percorsi operativi dal loro work-

flow abituale per la descrizione e l'archiviazione delle pubblicazioni. Al tempo stesso favorisce la disseminazione dei metadati grazie al protocollo OAI-PMH e fornisce un'interfaccia di ricerca semplice ed intuitiva che consente la rapida consultazione dei contenuti.

Riferimenti bibliografici

Cotoneschi, Patrizia. «Firenze University Press fra distribuzione tradizionale e libero accesso». *L'archivio E-Prints dell'università di Firenze: prospettive locali e nazionali. Atti del Convegno*. A cura di Patrizia Cotoneschi. Firenze University Press, 2004. 13–22. (Cit. a p. 154).

Gabaldo, Maria. «L'Open Archive e il suo sviluppo nell'Università di Verona». *Il Documento Digitale* I/MMXIII. (2013): 38–39. (Cit. a p. 161).

Maraviglia, M. Giulia. «L'Archivio E-Prints nella Biblioteca digitale dell'Università di Firenze». *L'archivio E-Prints dell'università di Firenze: prospettive locali e nazionali. Atti del Convegno*. A cura di Patrizia Cotoneschi. Firenze University Press, 2004. 5–12. (Cit. a p. 154).

Pasqui, Valdo. «Archivi di documenti elettronici: un modello di riferimento per la realizzazione della biblioteca ibrida». *Biblioteche Oggi* 18.10. (2000): 6–12. (Cit. a p. 152).

Pepeu, Giancarlo. «Introduzione». *L'archivio E-Prints dell'università di Firenze: prospettive locali e nazionali. Atti del Convegno*. A cura di Patrizia Cotoneschi. Firenze University Press, 2004. 1–3. (Cit. a p. 154).

GAIA INNOCENTI, SIAF–Sistema Informatico dell'Ateneo Fiorentino–Università di Firenze.
gaia.innocenti@unifi.it

CRISTINA MUGNAI, SIAF–Sistema Informatico dell'Ateneo Fiorentino–Università di Firenze.
cristina.mugnai@unifi.it

VALDO PASQUI, SIAF–Sistema Informatico dell'Ateneo Fiorentino–Università di Firenze.
valdo.pasqui@unifi.it

Innocenti, G., C. Mugnai. V., V. Pasqui. "FLORe – Florence Repository. L'Archivio ad accesso aperto dell'Università di Firenze". *JLIS.it*. Vol. 5, n. 1 (Gennaio/January

2014): Art: #9102. DOI: 10.4403/jlis.it-9102. Web.

ABSTRACT: The University of Florence has a long and consolidated experience in the exploitation of open archives and self archiving by researchers. After a short introduction on current open access context in Europe and in Italy, the article recalls past experiences by tracing the history of the first e-prints archive and then describes the genesis of FLORe, the new research institutional repository developed to promote open access archiving by teachers and researchers at the University of Florence. Afterwards the architecture and the main functional and technical features of FLORe are discussed aiming to depict and to motivate the approach followed in implementing FLORe as an extension of U-GOV Research, the catalogue of scientific products, rather than an autonomous repository. Following and implementing the requirements stated by the University of Florence Open Access Commission, FLORe proves to be the right solution because it reduces operational costs, does not add a new graphical user interface, preserves the already well consolidated end-users self-archiving workflow and, at the same time, provides a simple and intuitive search interface with OAI-PMH support for research metadata dissemination.

KEYWORDS: E-research; Institutional repository; Metadata; Open access; Open archive; Research data.

Submission: 2013-10-08
Accettazione: 2013-11-18
Pubblicazione: 2014-01-01